

Current Biology, Volume 31

Supplemental Information

Explaining face representation in the primate brain using different computational models

Le Chang, Bernhard Egger, Thomas Vetter, and Doris Y. Tsao

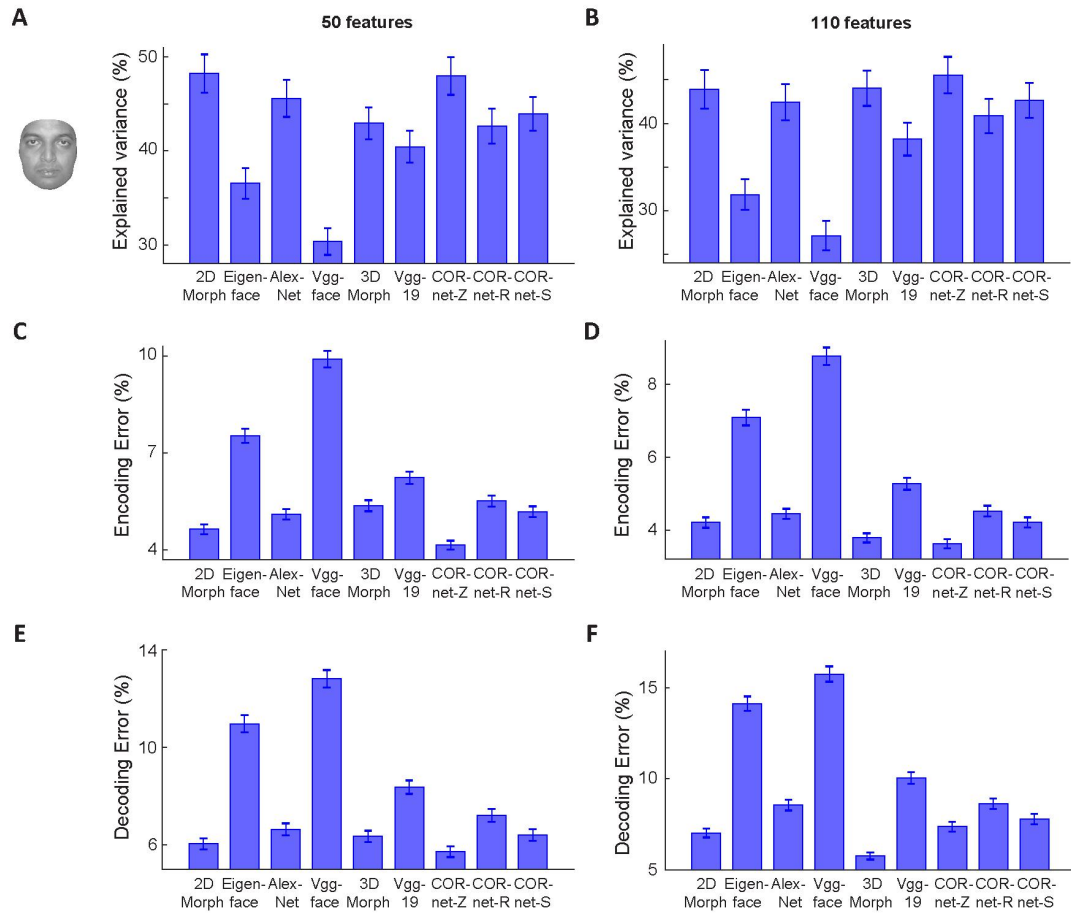


Figure S1. Encoding and decoding analyses using an alternative method of hair removal. Related to Figures 2 and 3.

After fitting the original images with the 3D Morphable Model, a mask was created using the fit, and the original image was cropped using the mask. A. Explained variances for 50 features of 9 different models using the cropped image as input. B, same as A, but for 110 features. C and D, same as A and B, but for the encoding errors. E and F, same as A and B, but for the decoding errors.

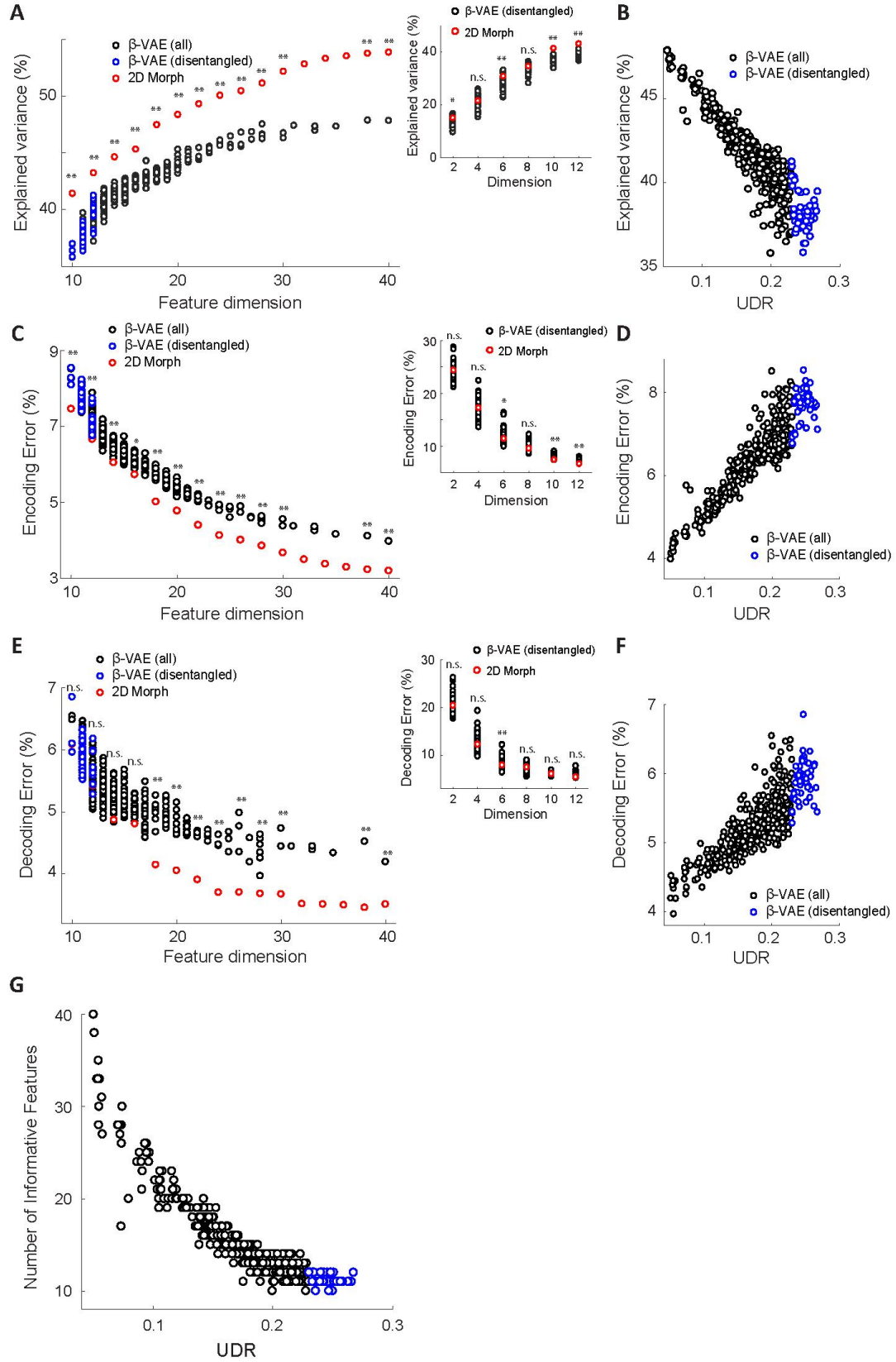


Figure S2. Comparing β -VAEs with 2D Morphable Model at equivalent dimensions. Related to Figures 2 and 3.

A, Explained variances for 400 β -VAEs after removing dimensions with variance < 0.01

were compared with the 2D Morphable Model at equivalent dimensions (equal number of shape and appearance dimensions were chosen for the 2D Morphable model). Wilcoxon signed-rank test was employed to compare the two models after performing 50-fold cross validation (*= $p < 0.05$; **= $p < 0.01$; n.s.=not significant). Inset, for the 51 most disentangled VAEs, subsets of features explaining the most variance of each model were compared to the 2D Morphable Model at equivalent dimensions (since equal numbers of shape and appearance dimensions were selected for the 2D Morphable model, only even numbers of total dimensions are shown here). B. Explained variances for all 400 β -VAEs are plotted against UDR score. C and D, same as A and B, but for encoding errors. E and F, same as A and B, but for decoding errors. G, Relationship between the number of informative features (variance ≥ 0.01) and UDR score for all 400 β -VAEs.

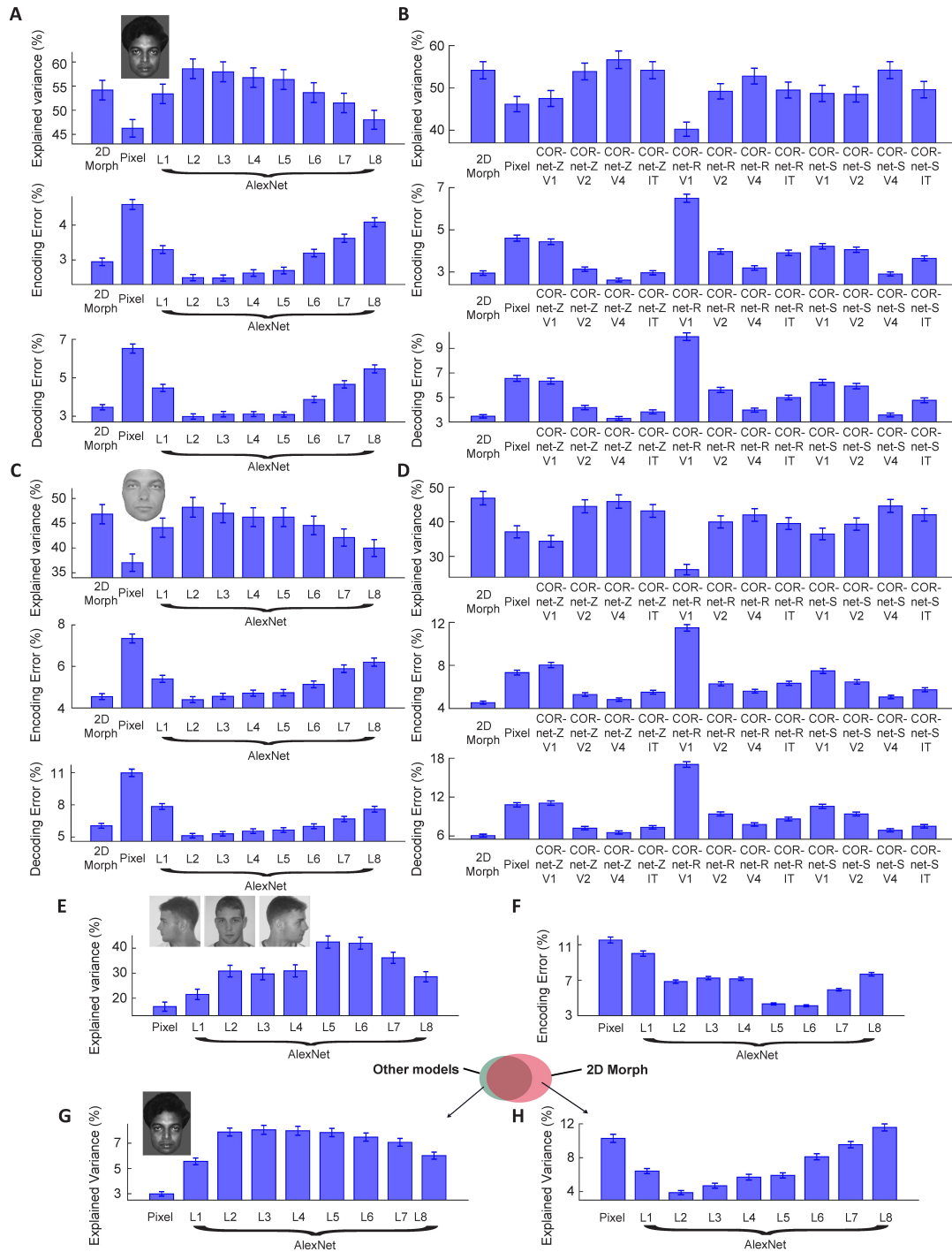


Figure S3. Encoding and decoding performance across different layers of AlexNet and CORnets. Related to Figures 2-4.

Inspired by previous works [26, 27], units in the penultimate layer were used to represent different neural network models, we finally examined how each individual layer of AlexNet and CORnets was related to neural responses. Quite surprisingly, we found intermediate layers of those models performed the best, and in particular, 2nd layer (L2) of AlexNet and V4 layer of CORnet-Z even outperformed the 2D Morphable Model (Panels A-D). We next asked whether the same conclusion could be

generalized to more complex stimulus sets: previously it was shown that only in face patch AM, neurons achieve full invariance to head orientations--a property unlikely to exist in early layers of CNNs due to the simpler feature selectivity of these layers. We examined the relationship between different layers of AlexNet and AM responses using a stimulus set containing facial images at 8 head orientations (Panels E and F). We found in this case, AlexNet L2 lost its advantage, with L5/6 performing the best. Given that AlexNet L2 cannot explain neural responses to the stimulus set with different head orientations, it is clearly not a viable candidate for explaining face patch activity. Finally, we asked how much unique variance each layer of AlexNet explained compared to 2D Morph by repeating the analysis in Figure 4A on individual layers of AlexNet, and found that the amount of unique variance of AlexNet L2 is similar to that of other intermediate layers (L3 to L6, Panel G). We also asked the converse, how much unique variance is explained by 2D Morph compared to each of the layers of AlexNet (Panel H). Here, we found what makes L2 distinct: the 2D Morph Model explains the least unique variance when compared to L2 activation.

A, Explained variances, encoding errors and decoding errors of the 2D Morphable Model and multiple layers of AlexNet for images after background removal (cf. Figure 2B). Pixel=input to the network. B, Same as A, but for CORnets. C and D, same as A and B, but for images fit by the 3D Morphable Model. E and F, Explained variances and encoding errors of a stimulus set containing 28 identities with 6-8 head orientations for multiple layers of AlexNet. Due to the small size of the stimulus set (n=220), 15 features were used instead of 50 to avoid overfitting, but the trend remained the same with 50 features. G and H, same as Figure 4A and B, but for multiple layers of AlexNet.

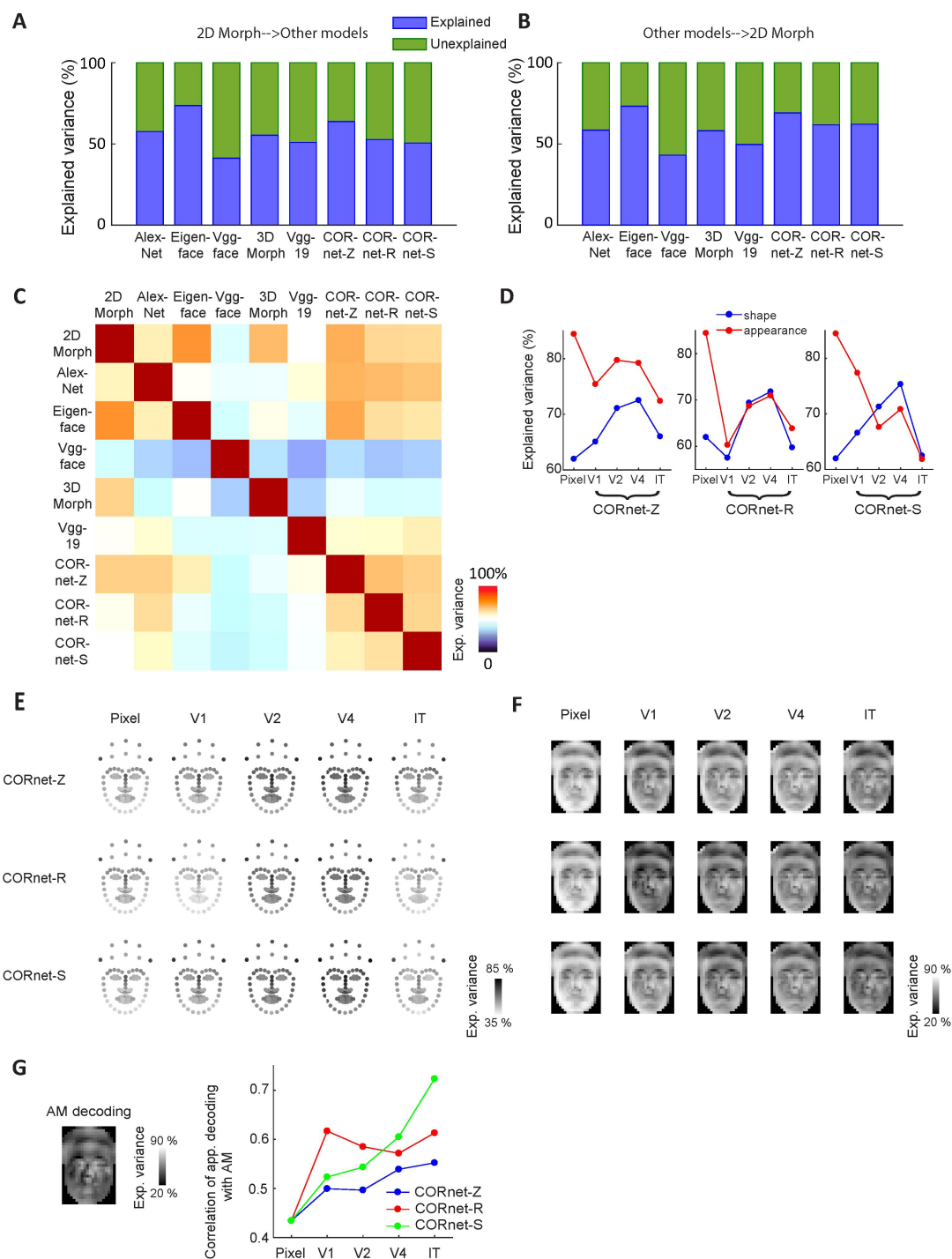


Figure S4. Direct comparison between feature spaces spanned by different models. Related to Figure 4.

A, Quantification of how much variance of features in one model could be explained by features of the 2D Morphable Model (similar to Figure 4B, but using feature values directly instead of neural responses). B, Same as A, but with features of the 2D Morphable Model being fitted by the other models. C, For each pair of two different models (X,Y), features of model X were fit by features of model Y, both using 50 feature dimensions (PCs). Explained variances were then averaged across the 50

PCs of model X, weighted by the variance of the original features explained by each PC. The averaged explained variances of all model pairs were then color-coded and plotted as a matrix, with its rows representing model Xs, and columns representing model Ys. D, Shape and shape-free appearance features for the 2D Morphabale Model were separately fit by 5 different layers (“Pixel”=input to the network) of three CORnet models. The V4 layer performed best for shape features, while the input layer performed best for shape-free appearance features. E, Decoding performance of all landmarks for different layers of CORnets. F, Decoding performance of shape-appearance descriptors for different layers of CORnets. All images were first morphed to the average shape. Pixel intensities of the shape-free image were averaged within multiple grids across the image (29*20 grids in total), and then fit with model features. G, Correlation of the decoding maps between CORnets and neural data.

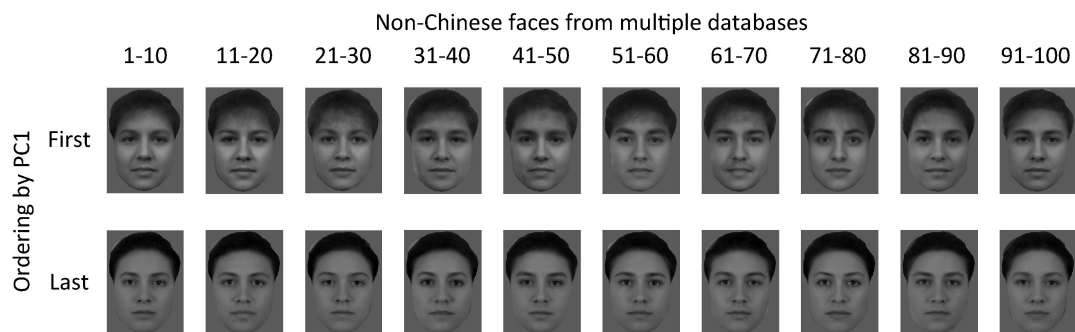


Figure S5. Comparison between VGG-face and AlexNet for non-Chinese faces. Related to Figure 5.

Same as Figure 5B, but for 748 non-Chinese faces. To attenuate the influence of diverse image backgrounds in multiple databases, we removed the background before presenting images to the networks (cf. Figure 2B).